Yeast

Yeast: One cell, one reference sequence, many genomes?

Erika Amethyst Szymanski, Niki Vermeulen, and Mark Wong

Science, Technology, and Innovation Studies; University of Edinburgh
Chisholm House, High School Yards
Edinburgh, EH1 1LZ

Published online in *New Genetics and Society* on 14 Oct 2019
doi.org/10.1080/14636778.2019.1677150

Abstract

The genome of *Saccharomyces cerevisiae*—brewer's or baker's yeast—was the first eukaryotic genome to be sequenced in 1996. The identity of that yeast genome has been not just a product of sequencing, but also of its use after sequencing and particularly of its mobilization in scientific literature. We ask "what is the yeast genome?" as an empirical question by investigating "the yeast genome" as a discursive entity. Analyzing publications that followed sequencing points to several "yeast genomes" existing side-by-side: genomes as physical molecules, digital texts, and a historic event. Resolving this unified-yet-multiple "genome" helps make sense of contemporary developments in yeast genomics such as the synthetic yeast project, in which apparently "the same" genome occupies multiple roles and locations, and points to the utility of examining specific non-human genomes independent of the Human Genome Project.

Keywords: genomics, yeast, discourse analysis

Word count: 8246

Introduction

What is a genome? For all that talk about genomes pervades contemporary life, what the term means is often unclear—not least because "genome" so clearly has multiple meanings. In social, philosophical, and historical studies of science, definitional discussions often focus on the tension between genomes as informational texts and genomes as physical molecules (e.g. Barnes and Dupré 2008; Kay 1998, 2000; Parry and Dupré 2010), genomes as genes versus genomes as DNA (Keller 2011), or the genesis of genomes as texts through genome sequencing. Those discussions also often focus on the human genome and the Human Genome Project (HGP) (e.g. Cook-Deegan 1996; García-Sancho 2006; Sapp 2003). This article has two central arguments: first, that genomes as sequences are constructed through how they are mobilized after sequencing as well as through how the sequence is produced; second, that, it is useful to consider the constitution of specific non-human genomes independent of the human genome.

The genome in which we take an interest is the reference genome for *Saccharomyces cerevisiae*, the first eukaryotic genome to be fully sequenced in 1996 and a widely used model for eukaryotic genomics. *S. cerevisiae*, brewer's or baker's yeast, uniquely bridges single-celled microorganisms and so-called "higher organisms;" yeast can be manipulated with molecular biology techniques as easily as the common lab bacteria *E. coli*, but unlike bacteria, yeast cell biology is understood to be fundamentally similar to that of humans and other metazoans. Though the single-celled fungi known as yeasts include hundreds of species, colloquially and in experimental biology "yeast" usually refers to this incredibly useful organism.[i] We examine publications in the years immediately following publication of the first yeast genome reference sequence, asking how researchers continuing to sequence yeast made use of "the yeast genome" as a rhetorical resource. In investigating the discursive construction of "the yeast genome," we aim to support the utility of investigating the meaning and manifestation of a particular genome as an empirical question.

The yeast genome project (1989-1996) coordinated sequencing efforts across North America (US and Canada), Japan, and the European Commission to yield a single unified reference genome sequence that could be seen as an "international achievement" (Levy 1994). From its initial release in 1996 through to the present, the Saccharomyces Genome Database (SGD) at the Stanford Genome Technology Center has hosted, curated, and provided free access to the resulting reference sequence and related resources. That reference, with relatively minor updates in the intervening 22 years, has remained in use and been considered of high quality into the next-generation sequencing era (Engel et al. 2013). This persistence is especially notable as the sequence coordinated the work of 94 laboratories across 19 countries, using technologies from manual Maxam-Gilbert and Sanger sequencing to automated Applied Biosystems sequencers (Goffeau et al. 1996).

We first outline the rationale for theorizing genomes as emerging in practice, with particular attention to textual practices, and for addressing what genomes are in terms of how "genomes" are mobilized in text. We then briefly review the history of the yeast genome before introducing our corpus analysis. We conclude by reflecting on some ramifications of the "yeast genome" for ongoing work in yeast genomics, including the synthetic yeast project.

Theorizing the yeast genome as a discursive entity and a rhetorical resource

Model organisms have become recognized as significant apparatuses and infrastructures for genetics and genomics, with scholarship tracing their development as laboratory technologies (e.g. Kohler 1994, Landecker 2007) and their roles in organizing research communities, databases, and associated working practices (e.g. Leonelli 2016, Nelson 2018). The characteristics of organisms, we

have seen, are inseparable from the epistemic practices and social contingencies of research through which they are constituted, and from biological knowledge produced with them. Less attention has been given to how model organisms take shape in the textual practices that comprise the circulating lifeblood of scientific knowledge. Scientific knowledge—as knowledge able to be shared and ratified by a scientific community—is primarily constituted in text, and particularly in peer-reviewed scientific articles (Bazerman 1984, 2000; Myers 1990). These texts constitute a significant social space in which scientists exchange resources and practice relationships through which science is produced. Indeed, many non-specialist readers will know the yeast genome only through reading, as it is instantiated in text; for the rest, discursive practices necessarily are framing devices or "terministic screens" (Burke 1968) through which physical work with genomes is understood. Research into "genomic discourse" often orients toward public understanding of science (e.g. Nerlich and Hellstein 2007), but social and political consequences also follow from how genomic discourse operates within scientific communities.

How scientific objects—including model organisms and the genomes through which they are so often employed—take shape through textual practice is thus just as important to understanding the characteristics of model organisms and how they shape biology as is understanding how they take shape in the lab, the stock center, and the database. The central argument of this paper is that the discursive construction of the yeast genome, or how "the yeast genome" is mobilized as a rhetorical resource, constitutes an important set of practices through which the genome takes and changes shape over time, with scientific texts an important (and too often sidelined) site in which organisms, communities, knowledge, and relations are co-constituted. That argument rests on pairing material-semiotics—through which genomes and other things are seen as temporarily stabilized assemblages of practices such that they become different through being practiced differently (Law and Lien 2012; Mol 2002)—with critical discourse modes of seeing of texts as social spaces wherein discursive practices enact scientific things (Fairclough 1995; Myers 1996).

What "genome" means has been a significant question as genomes, often ambiguously defined, populate an ever-expanding conceptual terrain. Genomes are physical molecules typically located inside cells. Genomes are also texts. Strictly equating or aligning the two raises epistemological and practical incongruities. "Genome" is also a conceptual tool that does not necessarily map directly to physical molecules or informational texts. Evelyn Fox Keller (2011) identifies four distinct meanings that have remained active, aggregating rather than excluding one another, since the term emerged in 1920: *genome-1*, the full complement of genes in a cell; *genome-2*, the full complement of chromosomes; *genome-3*, the full complement of DNA; and *genome-4*, the full complement of genetic material. All refer either to conceptual tools for doing classical genetics or to physical molecules existing prior to the explosion in sequencing activity, inviting the possibility of new additional meanings emerging and continuing to aggregate in the "postgenomic" era wherein genomes are most often understood in terms of sequence data (Keller 2015; Richardson and Stevens 2015).

One strategy for dealing with such multiple and potentially conflicting definitions is to choose one to underpin further discussion. Barnes and Dupré (2008) choose Brown's "definition of genomes as material" because it is in common use and, as they say, "nothing empirical follows from how terms are defined and users may define 'genome' as they wish" (78). Indeed they may, but it hardly follows that such choices of definition have no empirical consequences. Keller's principal point, in pointing to varied "genomes" side-by-side in the scientific literature, is to illuminate one such consequence—namely, that terms function as epistemic vehicles that "carry past conceptual frameworks into the present and future, sustaining the expectations and formulations that had originally generated these earlier meanings and infusing them into interpretations of new data"

(2011, 133). The persistence of those vehicles, she argues, ultimately hinders integrating new findings about genome function into revised biological theories. Her second point is to argue for her own "reformulation," the "reactive genome," in light of ongoing indications that genomes are more responsive organs than directors of cell activity. Elsewhere, she also leverages her analysis to argue against the application of Kuhn's assumptions about "overlapping kind-terms" to biology; while Kuhn proposes that overlaps presage either the dominance of one kind-term to the exclusion of others or forecast a split in discourse communities, both "gene" and "genome" have long been used in multiple ways (Keller 2012).

Lily Kay (1995, 1998, 2000), in tracing how genomes as "the book of life" became "informational and scriptural representations of life and heredity" (1995, 609), is in contrast interested in texts as *representations* and in the fundamental *misrepresentation* of DNA as a code. While some yeast sequences in GenBank and other databases might be understood as representations of living yeast cells, we will show below that the reference genome as a digital text does not simply represent a physical genome inside a cell. Moreover, because we construe genomes as realized through practice—discursive practices, in this analysis—texts are not representations but practices through which "the genome" continues to be constituted. This perspective allows us to examine what genomes become and how they can be mobilized rather than what they are in any essentialized sense.

Understanding "the yeast genome" as an accomplishment of and in practice sees that genome as assembled through ongoing use in addition to its initial sequencing. This analytical perspective offers at least two advantages. First, it affords a more satisfactory resolution to the problem of varied definitions: rather than choosing one as more fundamental, valid, or popular than the alternatives, all definitions can be understood as contributing to what is meant by "genome" as it aggregates meanings and is employed to do varied things in varied spaces of practice. Second, examining its construction in practice enables understanding the genome as not only physical and conceptual but discursive, as a rhetorical resource that can be drawn upon in texts as an important set of spaces where science happens.

A brief history of the yeast genome

Yeast is peculiarly absent in recent historical and social science literatures concerning model organisms in biological sciences—peculiarly, because yeast is a cornerstone amongst canonical model organisms for biomedical research.[ii] Yeast initially entered the laboratory as a function of longstanding roles in fermenting bread, beer, and wine, first indirectly through studies of Lavoisier and other chemists on fermentation and then famously in Pasteur's work for the wine and beer industries. Thereafter, yeast's ready availability, tractability, and industrial applicability made it an easy target for classical genetic work beginning in the 1930s. Later, in the 1960s and 1970s, yeast as a single-celled eukaryote enabled connecting the molecular genetics developed with phage and bacteria to biomedical studies of multicellular organisms. By the 1980s, yeast was well established as a "universal cell" (Herskowitz 1985), the simplest eukaryote, and a model for human biology—an obvious choice when complete eukaryotic genome sequencing became a reasonable scientific target.

Yeast genomes were invoked in two different ways by the early 20th century researchers generally considered responsible for establishing yeast genetics.[iii] Øjvind Winge, an applied scientist at Carlsberg Laboratory in Copenhagen, is routinely credited as the first to describe yeast sexual mating (Barnett 2011; Langer 2016; Szybalski 2001; Thorne 1947).[iv] Winge was thus responsible for early yeast genome studies in the sense employed amongst botanists of the same era, who employed "genome" to indicate the "haploid number of chromosomes" and conducted "genome

analysis" to characterize alterations in ploidy (Noguera-Solano, Ruiz-Gutierrez, and Rodriguez-Caso 2013, 215). Simultaneously, in the United States, Carl and Gertrude Lindegren were beginning to map yeast chromosomes. The Lindegrens, having studied in T.H. Morgan's laboratory where fruit fly chromosome mapping had gained so much traction,[v] saw mapping as the surest way of adapting yeast for classical genetics.

While the Lindegrens pursued increasingly outré (and unreplicable) theories of "gene action," Robert Mortimer at the University of California at Berkeley pursued increasingly detailed genetic maps, initially to understand why higher ploidy numbers appeared to increase yeast cells' sensitivity to radiation (Langer 2016). Mortimer's first map, published in 1960 with Donald Hawthorne from Roman Herschel's group at the University of Washington, located 22 genes in ten linkage groups. At a meeting of yeast geneticists the same year—later recast as the first international yeast genetics meeting ("The Yeast Genetics Community")—Mortimer offered to freely distribute the map along with the particularly amenable laboratory "wild type" strain used to generate it. Both came to define yeast genomics as it developed through the rest of the 20th century. The strain, S288C, became widely used by the international yeast genetics community and the strain sequenced in the yeast genome project.[vi]

Mortimer published twelve updates of the genetic map between 1960 and 1995, with later editions including markers gathered and solicited from across the yeast community. Many came from North America and western Europe, but his global network of contributors stretched as far afield as Moscow and the Shanghai Institute of Cell Biology (Mortimer technical records; Zhu and Kuang 1992). Though not every lab employed the same strain, these unified maps worked alongside the widespread popularity of S288C to gather and align an international community around "the exact same cell" (Boone 2014). Mortimer's lab coordinated this alignment by maintaining the authoritative versions of the community's information resources and, simultaneously, its physical resources. On the back of his offer to share the S288C strain, and as a by-product of mapping, Mortimer came to maintain the official-unofficial, universally recognized but never formalized yeast strain collection (Mortimer technical records). Having charge of both the Yeast Genetics Stock Center, as it came to be called, and periodic updates of a collective map, Mortimer became the central node of the international yeast genetics network. As a central point of exchange, he had the power to unofficially-officially enforce coordination between those informational and physical resources to ensure coherence around "the exact same cell."

In 1986, genetic "Mortimer maps" were joined by physical maps generated by Maynard Olson at Washington University St. Louis. Olson has credited the feasibility of physical mapping in the first place to the spread of S288C, together with what had by then become a community tradition of sharing clones and data (Olson, personal communication with G. Parolini). Understanding the community to be working on the same cell and genome also enabled Mortimer's and Olson's maps to be aligned and displayed together (Langer 2016). Similarly, both genetic and physical maps would eventually be merged with the reference genome sequence, enabling each to be understood as referencing the same "yeast genome." Later, Mortimer's maps were positioned as direct precursors of yeast genome sequencing and even as part of the same "project" to comprehensively describe and understand the yeast cell. When Mortimer and André Goffeau, the latter as coordinator of the European Community (EC)-led yeast sequencing project, were jointly awarded the 2002 Beadle Medal from the Genetics Society of America for career achievement, the citation credited Goffeau with having "brought to fruition" the project Mortimer started (Johnston 2003).

Yeast genome sequencing happened via an international project coordinated under EC funding and involving contributions from researchers in Canada, the United States, and Japan who

had initially been working independently. For the EC, yeast genome sequencing became an ideal solution to the perceived problem that Europe was failing to compete with American and Japanese biotechnology, in no small part because Goffeau was simultaneously a leading yeast biologist at the Université Catholique du Louvain and a civil servant in the EC Directorate for Science, Research, and Development (Goujon 2001). Goffeau was thus ideally positioned to propose yeast genome sequencing as a coordinated European project—a "capacity-building" exercise (Parolini 2018) to support nascent European sequencing companies and strengthen pan-European cooperation. Toward that end, Goffeau suggested distributing the work across a large number of existing laboratories rather than concentrating it in purpose-built sequencing centers. Yeast was uniquely suitable in this way because Europe already harboured numerous yeast laboratories whose cooperation could be compelled at least in part on the basis of the yeast community's self-consciously collaborative ethos.

Meanwhile, yeast had been named as one of several model organisms to be sequenced alongside and in advance of the human genome in the American NIH-funded HGP. Sequencing yeast and several other model organisms, it was argued, would develop new sequencing technology more cheaply and rapidly than tackling the human genome directly, while also providing presumably analogous genomes to aid interpretation of the human sequence once the latter had been obtained (Ankeny 2001; Cook-Deegan 1996; Mortimer technical records). James Watson, as then-leader of the National Center for Human Genome Research (NCHGR), offered to build a new sequencing center for David Botstein, then a vice-president of Genentech, and Ronald Davis, an established yeast geneticist at Stanford, to simultaneously sequence yeast and develop technologies toward human genome sequencing (Cherry interview 2018; Mortimer technical records). Botstein and Davis preferred to host those activities at Stanford, leading to construction of the Stanford Genetic Technology Center (SGTC).[vii]

Botstein and Davis's slowness to produce yeast sequence data while focusing on developing sequencing technology risked ill-will from the European project leaders. Mark Johnston at Washington University St. Louis (WUSL) joined the project to reassure the European partners with a more immediate American contribution (Johnston interview 2018). Having immediate access to Olson's clone libraries (Olson having left for the University of Washington) and WUSL's substantial cache of automated sequencers, Johnston made rapid headway on two chromosomes the European project had yet to begin. In addition, Goffeau's plans included making use of a chromosome I sequence already largely completed by David Kaback at the New Jersey Medical School with Jack von Borstel and Howard Bussey at McGill University and University of Toronto (Kaback 2013). A Japanese group with substantial inroads into sequencing chromosome VI, thanks to efforts to develop a state-of-the-art "super-sequencer" center at RIKEN, Japan's largest biology research institute, was also incorporated into the European-coordinated project (Cherry interview 2018; Langer 2016). Botstein and Davis eventually completed chromosome V with HGP funding.

In addition to chromosome III, published in 1992, complete sequences of chromosomes XI (1994), VIII (1994), II (1994), I (1995), VI (1995), and X (1996) were individually announced in *Nature, Science, EMBO*, and *PNAS*. The remainder, chromosome by chromosome, were published in a May 1997 supplement to *Nature*. The complete genome sequence, however, had been available online at the SGD since the previous year. The SGD has continuously supported and updated the reference genome and a growing catalogue of associated resources (now including many additional reference genomes of other *S. cerevisiae* strains) ever since. When S288C was resequenced from a single clone with next-generation sequencing technology in 2010, the authors of that study—the same group in charge of the SGD—found remarkably few differences between their sequence and the 14 year-old version cobbled together amongst diverse labs, libraries, and sequencing techniques (Engel et al.

2013). Though the original S288C reference sequence has been joined by genome sequences for a dozen other "wild types" and several hundred strains relevant to industry and studies in fungal ecology, the S288C sequence continues to scaffold contemporary work. As we will show below in terms of the synthetic yeast project, the stability and community assembly of the S288C reference sequence enables contemporary yeast researchers to construct material and historical ties to the previous "yeast community" to position their own work as a natural extension of what came before, and even part of the same project to comprehensively understand the yeast cell.

Beyond providing some historical background, this story of how the yeast genome has taken shape from the 1930s forward is important to supporting the analysis we undertake below. Mapping and sequencing the yeast genome assembled the work of many geographically and intellectually dispersed laboratories into a single information resource, building a community and a common understanding that the community was working on fundamentally the same thing and even on the same project. As a map and as a sequence, the yeast genome became singular in a way that united diverse physical manifestations of that genome into a common informational resource. This history provides strong reasons for understanding manifestations of "the yeast genome" in the scientific literature as cohering with a unified community idea of the yeast genome, and as seeing the discursive enactment as one set of practices through which the yeast genome has been enacted together with sequencing itself. Below, we examine how "the yeast genome" continued to be shaped through its discursive enactment after the sequence was released.

Method

Corpus construction

This study is part of a larger project (Medical Translation in the History of Modern Genomics, abbreviated TRANSGENE) concerning the organization of genome sequencing in the late twentieth century. That project has involved creating a bibliometric dataset capturing all new sequence submissions made to the European Nucleotide Archive for yeast, human, and pig between 1983 and 2000 (for yeast; dates vary slightly for each organism), together with publications linked to the first description of those sequences in the peer-reviewed scientific literature. The corpus is therefore essentially comprehensive of yeast sequencing work published after the reference genome was released—a set of papers almost certain to mention that reference. While a similarly constructed corpus for the human genome is much too large to permit detailed manual analysis, the comparatively small yeast literature is such that every mention of the yeast genome in the corpus could be analyzed manually. Another advantage of the corpus is that, while reflecting only a single category of use and users—the yeast genome was of course used by many scientists not conducting sequencing, and by industry members whose work did not necessarily result in publications (see, for example, Korhola 2018)—that focus permits an analysis virtually agnostic to nations, institutions, people, publication venues, funding, and research questions. Consequently, we can investigate how "the yeast genome" was mobilized by yeast sequencers without limitation to any one national or institutional context.

This analysis began with a data set capturing all publications identified in SCOPUS as linked to the first description in the peer-reviewed scientific literature of a new sequence submitted to the European Nucleotide Archive (ENA) between 1983-2000, inclusive. The ENA, the US-based GenBank, and the DNA Data Bank of Japan are synchronized daily per the International Nucleotide Sequence Database Collaboration such that each contains the same complement of sequencing reads, alignments, assemblies, and annotations (http://www.insdc.org; see also Stevens 2018). With each new submission to the ENA, submitters are asked to indicate any publications associated with the

submitted sequence. Publications can also be added after the initial submission. Our data set captures all such publications by PubMed ID number along with metadata including the submitter's name, institutional affiliation, location (country), and year of submission. The journal title in which the associated publication appeared and the publication date as reported by the journal were manually added for publications recorded in years 1996-2000.

Unsurprisingly, dates associated with sequence submissions in the ENA differed from dates of associated publications about 37% of the time—we expected that sequences would sometimes be submitted before publications referencing those sequences appeared. Surprisingly, however, publication dates as reported by journals sometimes *preceded* sequence submission dates reported by ENA. Following conversations with ENA staff, we hypothesize that sequence submitters may sometimes have misunderstood database protocols and associated a new sequence with a prior publication. Submitters may have also incorrectly created an entirely new accession number when trying to record an update to an existing sequence, which might then be associated with an earlier publication. Large updates greatly expanding an existing sequence are occasionally given a new accession number, which would yield a similar result.

These discrepancies have been handled by omitting publications with ENA dates between 1996-2000 but journal publication dates outside that range, and by accepting that some number of publications with ENA dates outside 1996-2000 but journal publication dates within that range have been missed by our protocol. These potential absences are made less concerning by finding coherence and saturation across our data such that analytical categories are very unlikely to change with the inclusion of additional papers. However, we cannot rule out the possibility that the elided set of publications is substantially different in some way relevant for our analysis. Our search strategy returned 171 articles. Of those 171 articles, 19 had publication dates prior to 1996 and six had publication dates after 2000, leaving a corpus of 158 articles for analysis.

Corpus analysis

Articles collected as described above constituted a corpus for textual analysis. Full texts were located in PubMed via PubMed ID number and obtained via PubMed Central or the publisher's website. Publication data obtained from bibliometric analysis (author, institution, journal, date of publication) was manually checked. As noted above, conflicting publication dates were identified in approximately 37% of cases. No other discrepancies were found.

For articles available as webpages or as optical character recognition-enabled pdfs, full texts were searched for "genom." Unsearchable articles were manually scanned for the same features. When instances of "genom" across a single publication consisted solely of "genomic" as a modifier in the context of genomic libraries, genomic DNA, or genomic fragments, data from that publication were coded with that phrase. When a publication mentioned "the genome," "the genome project," "the Saccharomyces genome database," "the yeast genome sequence," or other similar reference to "the genome," the relevant sentence(s) and surrounding context were examined and coded with respect to the function and location of "genome" in that context. Descriptive codes were developed organically, then refined to eliminate redundancy after all segments were coded; sentences were then re-examined through the refined code book. Titles and abstracts were scanned for the scientific contribution each publication aimed to make.

Of 158 articles retrieved by our search protocol, full texts were unavailable for three. Three more were excluded because they directly contributed to the genome project and our interest is in "the genome" after the reference sequence's release. 16 of the remaining 152 made no mention of

"genome," "genomes," or "genomic." 42 contained references to "genomic"—genomic DNA, genomic libraries, genomic fragments, etc.—without mentioning "genome" or "genomes." 78 articles mentioned "genome" or "genomes" in the context of "the yeast genome," "the S288C genome," "the yeast genome project," "the yeast genome data base," and related phrases.

No quantitative conclusions were drawn because while potential imperfections of the corpus noted above are very unlikely to alter qualitative analysis, even minor incompleteness could have significant effects for a quantitative analysis given the small numbers of articles involved. As no inferences were drawn from code frequency, the length of meaningful coded segments was established empirically in context. In the analysis below, all quotes are indicative of categories containing more coded segments than would be possible or desirable to provide in full.

Analysis: What is "the yeast genome?"

In this corpus, "how was the genome mobilized?" can be answered in two registers: in terms of *what the genome does* and *where the genome is*. The second is in many ways a consequence of the first; the "genome" is called upon for some function and, in so doing, one of multiple potential visions of what the genome is and where it resides is invoked. In the following examples, "genome" functions both as a noun, as in "the yeast genome," and as a modifier in compound nouns, as in "the genome project" or "the genome data base." All of these phrases invoke or call into being a "yeast genome" in that the genome project or genome data base imply a genome, and the project and database enact that genome in particular ways: as physical materials stored in cells, informational texts stored in digital databases, and a historic event shared by a scientific community. Articles are identified by PubMed identifier (PMID).

The genome is sometimes a source of physical experimental materials, especially in materials and methods sections of papers, where "genomic DNA," "genomic fragments," and "genomic libraries" frequently appear.

> "*The YND1 gene (GenBank™/EBI accession number U18778(30)) was cloned by PCR using S. cerevisiae genomic DNA as a template*." (PMID 10409709)
>
> "*The 5.4 kb-DNA fragment was cloned from the genomic library of the L-azetidine-2-carboxylic acid-resistant mutant*." (PMID 10894734)
>
> "*Standard methods were used for the introduction of DNA into yeast, the preparation of genomic DNA, and for tetrad dissection (Rose et al. 1990)*." (PMID 9832547)

These phrases are ambiguous about invoking the genome as a physical object itself or as a concept enabling or guiding the selection and acquisition of physical materials in the way that "gene" as a concept enabled locating associated physical structures without necessitating that those structures were understood to be co-equivalent with the gene itself—a necessary slippage to enable the genome as the cell's total collection of genes to become the genome as the cell's total collection of chromosomes (Keller 2011, 2015).

Relatedly, the genome is sometimes a physical context for manipulating other experimental materials, a landscape:

> …into which additional DNA can be integrated: "*The vti1-1 mutation was integrated into the genome for further analysis*." (PMID 9199167)
>
> …or able to be surveyed for features of interest: "*The null mutations were confirmed by PCR analysis of the yeast genome.*" (PMID 9753630)

"Genomic," in these cases, necessarily calls upon what Keller (2011) terms *genome-3*—"an organism's complete set of DNA" (133). Experimental technique enacts "complete" in different ways

depending, for example, on the method of preparing a "genomic library," of extracting "genomic DNA," or of selectively analyzing by PCR, though in all cases the genome comprises physical material inside yeast cells.

In contrast, in other instances, the genome is a set of instructions for making other molecules:

> "*While most yeast snoRNAs are independently transcribed from mono-, di-, or polycistronic units, the relatively compact* S. cerevisiae *genome does encode a few intronic snoRNAs, U24, U18, snR38, snR39, and snR59, processed according to the vertebrate mode.*" (PMID 9891049)
>
> "*The cDNA insert contained within the plasmid was cloned and was found to encode an open reading frame (ORF) from the yeast genome (YER143w; GenBank accession no. U18917 and AF034895*)." (PMID 10330187)
>
> "*The mouse eIF3-p44 sequence was used to conduct a TblastN search of the entire yeast genome sequence, and a putative protein was identified whose amino acid sequence exhibits 33.3% identity to mouse eIF3-p44.*" (PMID 10085088)

This "genome" is not physical material in the cell so much as information describing yeast biology. While the "genome" encoding snoRNAs (small nucleolar RNAs) in the first example may be a physical molecule, it is a molecule seen through the controlling metaphor DNA IS A CODE (to use Lakoff and Johnson's notation; 1980), making information content that molecule's defining property without insisting upon its physicality. In the second and third examples, the genome is not a physical molecule at all but a textual resource housed in GenBank with features defined in terms of reading frames. Here, the informational genome has moved from the cell to a digital database, from a physical molecule to a text.

If the digital genome is not assumed to solely a *representation* of a physical genome, we might instead see "the genome" as a discursive entity moving from the cell to the database but without ever removing "the genome" from the cell, aggregating rather than excluding. This move and its inherent multiplying has consequences for the shape of the community that continues to participate in constructing the genome as a digital text. It also has consequences for the work the genome can perform. As Soraya de Chadarevian (2004) observes for the *C. elegans* genome sequence, moving the genome from cell to database enables it to become a new kind of mapping tool:

> …enabling the chromosomal location of cloned genes to be determined: "*Once the complete sequence of the S. cerevisiae genome was available, we were able to locate the PMP3 gene on the right arm of chromosome IV (YDR276c), between genes YDR275w and MTH1 (YDR277c), from position 1 013 467 to position 1 013 634*." (PMID 10835350)
>
> …permitting observations about genome configuration to support a hypothesis: "*Intriguingly, in the yeast genome the U14-coding sequence is close to that for another box C/D snoRNA, snR190, which suggests that both snoRNAs might be produced by processing of a common dicistronic transcript*." (PMID 9891049)
>
> …allowing one gene sequence to be used to identify potential homologs: "*In the genome of S. cerevisiae S288C, there are seven open reading frame sequences sharing homology with the FLO1 gene.*" (PMID 9851992)
>
> …or, conversely, the absence of homologs: "*A survey of the recently released complete S. cerevisiae genome with ALDP, PMP70, and Pxalp sequences failed to detect additional homologs with smallest sum probability scores*." (PMID 8876235)

Consequently, the informational genome can be used to manufacture coherence across knowledge generated by disparate labs by making it possible to line up textual results—often quite literally, in the case of genome sequences as alphabetic texts—and read across them. This, as Latour and Woolgar (1986) generalize in *Laboratory Life*, is the fundamental scientific process—flattening

individual, situated, three-dimensional instances into textual abstractions so that texts can be lined up in one place and time, seen as equivalent representations of a generalized phenomenon, and employed to support a conclusion that applies to and beyond all of them. Prior to the availability of the reference genome, such coherence across the yeast community was generated by strain sharing practices and standard protocols and notation for documenting variation from the reference strain, S288C, authoritatively embodied in living cells housed at Mortimer's stock center. After the sequencing project, part of that work was transferred to the informational, textual yeast genome, making the practice of lining up and making the same available to those who worked with the digital text as well as those who worked with the physical cell. Simultaneously, a new kind of equivalence could be manufactured, between the textual yeast genome and other textual genomes.

Thus, the complete genome sequence also enabled "mapping"[viii] relationships across species:

> …in terms of gene families: "*The nucleotide sequence of the complete open reading frame has been determined, and the deduced amino acid sequence has been compared with MBF1 sequences of other organisms. yMBF1 has 43% amino acid identity with the Drosophila counterpart*." (PMID 9710580)
>
> …to locate homologs for further analysis: "*Southern blot hybridization of EcoR1-, HindIII, or BamH1-digested genomic DNA indicated that there is a single gene for the factor within the Bombyx genome*." (PMID 9207077)
>
> …and in terms of evolutionary relationships explicitly: "*Since the genome projects of these organisms have been concluded (Goffeau et al., 1996; C. elegans Sequencing Consortium, 1998), along with many bacteria genome projects, it can be established that while no DSCR1-like genes are to be found in bacteria, they do appear in lower unicellular eukaryote organisms, suggesting an involvement in a function arising from the transition from prokaryotes to eukaryotes*." (PMID 10756093)
>
> …or implicitly: "*None of the human or yeast protein subunits of RNase P exhibits significant sequence similarities to any predicted ORFs [open reading frames] from eubacterial or archaeal genomes*." (PMID 9618478)

These genomes continue to be a text rather than a set of molecules, but the essential features of that text have changed from merely containing information to comprising complete, searchable information. Being both comprehensive with respect to the organisms they describe and completely searchable, they can be made functionally equivalent across species, aligned and compared—their varied organismal identities are functionally less important than their textual equivalence.

As a comprehensive text, the complete genome sequence enabled making claims about gene families or their absence, as exemplified above, but also about:

> … differences amongst strains: "*Unlike MUC1, the STA1 to -3 genes are not present in the genomes of the S288C-derived laboratory strains that were used in the sequencing of the S. cerevisiae genome*." (PMID 10515942)
>
> … the quality of newly generated sequence data: "*The nucleotide sequences from each end to 30th were confirmed identical to those registered in the yeast genome database, although some differences were observed at several positions*." (PMID 10873545)
>
> …or, rarely, corrections to the reference genome: "*While preparing expression constructs to investigate the biochemical properties of Fat1p, the product of the FAT1 gene, we detected an error in the sequence as contained in the Saccharomyces Genome Data Base*." (PMID 9660783)

Finally, the genome is sometimes the genome project as important scientific work requiring acknowledgment.

> "*YKL741 is an open reading frame (ORF) found by the yeast genome sequencing project*." (PMID 8876235)
>
> "*A nearly identical sequence derived from the left arm of chromosome XV was recently identified by the yeast genome sequencing project (Cassamayor 1995)*." (PMID 8978028)

At times, citing the genome project is less about using information from that project, more about signifying community membership and indicating the authority of the present authors to speak to genome-related questions by virtue of their awareness of previous work and intent to enter into the same conversation. This "genome" is neither physical molecule nor informational text, but a rhetorical device for enacting community and part of the scientific context in which subsequent science happens (Gilbert 1977). In several cases, "the genome project" is the only mention of "genome" in the paper, a *conceptual* resource mobilized to indicate participation in the long-term collective project of yeast genomics.

All of these "genomes"—a complement of physical molecules in yeast cells, information in a database, and a historical event and conceptual resource in the literature—exist across the entirety of the corpus and often coexist in the same publication. Genome sequencing—or, rather, the public availability of a complete reference genome sequence—could be seen as representing physical molecules inside yeast cells in a new way, creating a more detailed map of these molecules to add to genetic and physical maps. The S288C reference sequence, however, does not map onto any one set of physical molecules; it is a *new* articulation of what a genome might mean, not a representation of a previously existing one. The reference sequence did not replace previous yeast genomes but add to them—scientists continued to do things with maps and molecules. This multiplicity contrasts with "the exact same cell" at the center of the yeast community but necessarily depends on it, as researchers could only mobilize "the genome" in multiple ways after the physical, informational, and conceptual infrastructural work had been accomplished to ensure that they all contributed to a coherent body of knowledge about a scientific object, the yeast cell, held in common.

Because these genomes were all understood as essentially the same thing, the reference sequence changed the shape of "the yeast genome," along with the range of available options for working with it and the community to whom those options may be available (see Mol 2002 on the work of discursive assemblages). Any response to the question "what is the yeast genome?" needs to account for this unity-in-multiplicity to make sense of how the genome has been mobilized; simply choosing one to employ one of those facets to represent the whole, or assuming that the real genome is any one of the four different potential physical genomes Keller identified, is insufficient. Understanding what work the yeast genome can do, where it was and where it could go, requires seeing the multiplicity as well as the unity, as the contemporary synthetic yeast project demonstrates.

Synthesis: Making "the yeast genome" after the yeast genome sequencing project

Saccharomyces cerevisiae 2.0 or the synthetic yeast project exemplifies how the simultaneous unity and multiplicity of the yeast genome matters to the structure and organization of ongoing work with that genome. The synthetic yeast project is an international consortium of synthetic biologists aiming to build the first complete and comprehensively redesigned eukaryotic genome entirely from laboratory-synthesized DNA. Re-imagining the yeast genome reference sequence as *Saccharomyces 1.0* and positioning itself as a natural follow-on, the synthetic yeast project is, as its leader Jef Boeke

has phrased it, “upgrading yeast with a twenty-first century version of its operating system” (Szymanski 2018). Among other goals, the synthetic yeast project aims to complete what yeast genome mapping and sequencing promised: a comprehensive understanding of yeast as a basic eukaryotic cell. And like mapping and sequencing, the magnitude of DNA synthesis and assembly involved in the synthetic yeast project requires cooperation among geographically and intellectually distributed laboratories, each charged with assembling one or two of yeast's sixteen chromosomes (see syntheticyeast.org).

The synthetic yeast project has much in common with the yeast genome sequencing project. Both coordinate international consortia to create a single coherent resource; both aim to generate a new kind of yeast genome; both use yeast as the first eukaryote and a model for other eukaryotes for scaling up a strategy only previously applied to bacteria. Both work with fundamentally “the same” organism, as the template for the synthesis project is a close descendant of the S288C strain used for yeast genome sequencing (Richardson et al. 2017). The synthetic yeast project now depends on employing that genome simultaneously as a complete informational text housed in infinitely shareable digital databases, and as a set of physical molecules housed in yeast cells that can be physically manipulated and replaced, and as a historic accomplishment which the present consortium now seeks to recapitulate.

The genome was initially redesigned *in silico*, reducing the original S288C reference genome from a 12,071,297-character to a 11,352,534-character text via a set of algorithms to add, remove, and rewrite the features of that text (Richardson et al. 2017). This genome can be seen exclusively as a text—like the reference sequence, the redesigned Sc2.0 sequence is derived from a physical cell but does not represent it; nowhere is there a physical cell wherein that genomic text is instantiated as a set of molecules. That textual *in silico* genome is then used to guide the physical assembly of the genome—genomic DNA chemically synthesized without ever seeing the inside of a yeast cell or the rest of the genetic context that would ordinarily constitute it as “genomic” rather than simply “DNA.” That genomic DNA, replicated chemically by polymerase chain reaction (PCR) or biochemically in *E. coli* or yeast cells, is then linked together to form a “chunk” of a genome in a yeast cell used as an assembly tool (Szymanski and Calvert 2018). The chunk is a section of a genome which will only manifest as a genome when chunks are assembled through the collaborative work of a consortium which has already redesigned “the genome” (a text), has been working to construct “the genome” (physical molecules) for years, and will at some point produce the synthetic genome as a significant scientific event. “The genome,” a yeast genome seen to be fundamentally the same thing such that its instantiations can be treated as equivalent to each other, must at one and the same time be textual—residing in several different digital forms—physical—residing in cells where genomic molecules can be replaced by other molecules which become genomic in context—and historic—residing in the community such that the consortium’s work can be construed as “2.0” in extension and in response to a preceding version. Each of these genomes can be manipulated or called up individually without invoking change in the other; simultaneously, all must be understood as fundamentally unified and interchangeable for the project to construct a yeast genome by replacing a genome inside yeast cells with molecules built according to directions provided by a text inside a computer.

Seeing texts (and the peer-reviewed literature in particular) as a location where the genome is practiced also helps make sense of how the synthetic yeast project itself changes the shape of the yeast genome—not solely as a future physical molecule materially different compared to its parent, an informational text with a different set of letters, and a future scientific milestone, but also a discursive entity, “synthetic yeast,” that is already acting on science and broader society (Szymanski, forthcoming). Like the sequencing project before it, this synthesis project is being framed as a model,

prelude, and opportunity for technology development in advance of a human cognate. In light of the possibility of a synthetic human project, understanding the identity of the "genome," the relationships embedded within it, and relationships amongst "genomes" and cells or organisms clearly become more than technical problems.

Conclusions: The enduring "yeast genome"

This paper asks what is "the yeast genome?" in empirical terms of how "the yeast genome" was enacted after the reference genome sequence was released. Finding several "yeast genomes" juxtaposed in the yeast sequencing literature is hardly surprising given the multiplicity of definitions other authors have identified for "genomes" more generally (e.g. Keller 2011; Noguera-Solano, Ruiz-Gutierrez, and Rodriguez-Caso 2013). Articulating those genomes and how users mobilized them, however, has consequences for yeast's position in genomics and its continued mobilization in synthetic genomics.

The S288C genome scaffolds other yeast genome sequencing work: providing a reference against which other genomes are compared; a source of functional annotation information; and material, digital, and cultural infrastructure for collaborative projects. More broadly, yeast has functioned as a model in numerous respects: as a stable target for genetic modeling (Keller 2009), a tractable "preexplanatory" (Ankeny 2000) biochemical model for elucidating genotype-phenotype descriptions, a "universal cell" (Herskowitz 1985) able to stand in for human cells, a model eukaryotic organism (Langer 2016), a model or practice genome for the HGP, and an explanatory (Ankeny and Leonelli 2011) model genome with evolutionary correspondence to the genomes of humans and other metazoans. Understanding yeast as the simplest eukaryotic cell enabled targeting yeast in the first eukaryotic genome sequencing project; genome sequencing in turn enabled the community imaginary of "solving" the cell by characterizing all of its roughly 6,000 open reading frames. Imagining a potential "solution" to the cell might itself be seen to enable contemporary synthetic biology work toward making yeast a fully described, fully controllable cell factory and "chassis" for bioengineering. The "golden age of biology" (Johnston 2000) ushered in by the reference sequence was not only about geneticists' newfound power in comprehensively understanding the cell, but about cognate power extended to medicine via understanding human disease-related genes through yeast homologs, about improving industrial yeasts by understanding the genetic composition of their desirable qualities, and about redesigning yeast for other purposes.

Questions about relationships amongst various yeast genomes relate to how yeast as a physical resource became yeast as an informational resource, enabling yeast-based technologies such as yeast artificial chromosomes (YACs) and the synthetic yeast project that derive fungal identity chiefly from sequence information. Beyond their utility in understanding how research objects and thus scientific knowledge have been made—understandings with their own practical consequences—these questions have practical governance implications.

Genomics has fundamentally altered scientific and broader social understandings of organisms, including humans. Those understandings are tied to genomes' mobilization in discourse because discourse is how genomes as scientific entities travel through scientific and other societies. As yeast continues to function as a model organism for work with human cells, and as synthetic yeast becomes a preliminary exercise toward constructing a synthetic human genome—a move happening in the form of the Genome Project-Write (Boeke et al. 2016; Pollack 2016)—we will be impelled to ask: how does *Homo sapiens* 2.0 relate to human cells, and to whole human organisms?

Acknowledgements

This research was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 678757 awarded to Miguel García-Sancho for the project 'TRANSGENE: Medical translation in the history of modern genomics." We thank the members of TRANSGENE including Dr. García-Sancho, Ann Bruce, Rhodri Leng, James Lowe, Giuditta Parolini, and Gil Vry. We particularly thank Dr. Parolini for her generosity in sharing data and expertise.

[i] The major exception is the "yeast" of yeast infections, *Candida albicans*, which is evolutionarily distant from *S. cerevisiae*.

[ii] Most of the historical attention to yeast as a research organism has come from yeast biologists, notably including an edited collection of personal essays by leaders in the field published by Cold Spring Harbor Press in 1993 (ed Hall and Linder) and a monograph by British yeast taxonomist James Barnett reviewing scientific advances in yeast research from 18th century chemistry through the 20th century.

[iii] It is not clear whether or how Winkler's understanding of nuclear genetic material as the "genome" influenced either Winge's or the Lindegrens' approach to genetics. I use the term "genome" here to indicate the continuities between Winge's and the Lindegrens' descriptions of yeast genetics and later genome mapping work, but without intending to suggest when the "yeast genome" as a term or as a term used in any particular way first appeared.

[iv] Jan Satava theorized haploid-diploid alternation well before Winge, in 1918, but in Czech and thus without much international fanfare until his work was re-presented in conjunction with Winge's findings at a meeting in Amsterdam in 1935 (Barnett 2011).

[v] Carl and Gertrude Lindegren worked together during Carl's PhD and throughout his career, though only Carl was awarded the PhD and Gertrude was rarely named as a scientific contributor (see Langer 2016; see also the dedication in Lindegren 1949).

[vi] Though S288C is routinely named as the strain sequenced in the yeast genome project, and the reference sequence has consistently been described as "S288C", the sequencing project also involved several closely related strains including FY1679 and AB972 (Engel et al. 2013). That the differences amongst these variants seem to have been unimportant in assembling the reference sequence speaks to the sequence being understood as a reference rather than a representation of any specific laboratory yeast. It also speaks to the work done by the widespread adoption of S288C and strain sharing practices in the sense that, by the time the reference sequence was being organized, yeast geneticists could be confident that their local variations on the common S288C were essentially comparable.

[vii] The Stanford Genetic Technology Center (SGTC) was initially founded as the Stanford DNA Sequencing and Technology Center, but changed names before 1996 and was recorded at the conclusion of the yeast genome project as the SGTC. The SGTC persists today as the Stanford Genome Technology Center in a building near its original location. See http://med.stanford.edu/sgtc/general/history.html.

[viii] Scott F. Gilbert (2004) has articulated some of the incongruencies that follow from employing the "mapping" metaphor broadly across genetics, particularly from the perspective of evolutionary developmental biology. David Gugerli, in the same volume, advocates for the utility of understanding "mapping" as a "communicative strategy" rather than questioning or criticizing the limits of its appropriate metaphoric usage.

References

Aguilar, Alfredo, Torbjörn Ingemansson, Stéphane Hogan, and Etienne Magnien. 1998. "Industrial Platforms—a Unique Feature of the European Commission's Biotechnology R&D Programme." *Trends in Biotechnology* 16 (9): 365–68.

Ankeny, Rachel A. 2000. "Fashioning Descriptive Models in Biology: Of Worms and Wiring Diagrams." *Philosophy of Science* 67: S260–72.

Ankeny, Rachel A. 2001. "Model Organisms as Models: Understanding the 'Lingua Franca' of the Human Genome Project." *Philosophy of Science* 68 (3): S251–61.

Ankeny, Rachel A., and Sabina Leonelli. 2011. "What's so Special about Model Organisms?" *Studies in History and Philosophy of Science Part A*, Model-Based Representation in Scientific Practice, 42 (2): 313–23.

Barnes, Barry, and John Dupré. 2008. *Genomes and What to Make of Them.* Chicago: University of Chicago Press.

Barnett, James A., and Linda Barnett. 2011. *Yeast Research: A Historical Overview*. Washington, D.C.: American Society for Microbiology Press.

Bazerman, Charles. 1984. "Modern Evolution of the Experimental Report in Physics: Spectroscopic Articles in Physical Review, 1893-1980." *Social Studies of Science* 14 (2): 163–96.

Bazerman, Charles. 2000. *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. WAC Clearinghouse Landmark Publications in Writing Studies. WAC Clearinghouse.

Boeke, Jef D., George Church, Andrew Hessel, Nancy J. Kelley, Adam Arkin, Yizhi Cai, Rob Carlson, et al. 2016. "The Genome Project-Write." *Science* 353 (6295): 126–27.

Boone, Charles. 2014. "Yeast Systems Biology: Our Best Shot at Modeling a Cell." *Genetics* 198 (2): 435–37.

Burke, Kenneth. 1968. *Language as Symbolic Action*. Berkeley: University of California Press.

Cherry, J. Michael. 2018. Unpublished interview with Erika Szymanski. 16 April 2018. Palo Alto, California, USA.

Chadarevian, Soraya de. 2004. "Mapping the worm's genome: tools, networks, patronage." In *From molecular genetics to genomics: the mapping cultures of twentieth century genetics,* edited by Jean-Paul Gaudillière and Hans-Jörg Rheinberger, 95-110. London: Routledge.

Cooke-Deegan, Robert. 1996. *The Gene Wars: Science, Politics, and the Human Genome*. New York: W.W. Norton.

Dujon, Bernard. 2015. "Basic Principles of Yeast Genomics, a Personal Recollection." *FEMS Yeast Research* 15 (5).

Engel, Stacia R., Fred S. Dietrich, Dianna G. Fisk, Gail Binkley, Rama Balakrishnan, Maria C. Costanzo, Selina S. Dwight, et al. 2013. "The Reference Genome Sequence of Saccharomyces Cerevisiae: Then and Now." *G3: Genes|Genomes|Genetics* 4 (3): 389–98.

Engel, Stacia R, Marek S Skrzypek, Sage T Hellerstedt, Edith D Wong, Robert S Nash, Shuai Weng, Gail Binkley, Travis K Sheppard, Kalpana Karra, and J Michael Cherry. 2018. "Updated Regulation Curation Model at the Saccharomyces Genome Database." *Database: The Journal of Biological Databases and Curation.* doi10.1093/database/bay007

Fairclough, Norman. 1995. *Critical Discourse Analysis*. London: Routledge.

Fields, Stanley, and Mark Johnston. 2005. "Whither Model Organism Research?" *Science* 307 (5717): 1885–86.

García-Sancho, Miguel. 2006. "The Rise and Fall of the Idea of Genetic Information (1948-2006)." *Genomics, Society and Policy* 2 (3): 16–36.

Gilbert, G. Nigel. 1977. "Referencing as Persuasion." *Social Studies of Science* 7 (1):113–22.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, and et al. 1996. "Life with 6000 Genes." *Science; Washington* 274 (5287): 546–67.

Goujon, Phillipe. 2011. *From Biotechnology to Genomes: The Meaning of the Double Helix*. Singapore: World Scientific Publishing Company.

Hall, Michael, and Patrick Linder, eds. 2003. *The Early Days of Yeast Genetics*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Herskowitz, Ira. 1985. "Cell Biology: Yeast as the Universal Cell." *Nature* 316 (6030): 678–79.

Johnston, Mark. 2000. "The Yeast Genome: On the Road to the Golden Age." *Current Opinion in Genetics and Development* 10 (6): 617–23.

Johnston, Mark. 2003. "The 2002 George W. Beadle Medal Robert Mortimer and André Goffeau." *Genetics* 164 (2): 422–23.

Johnston, Mark. Unpublished interview with Erika Szymanski. 9 April 2018. Boulder, Colorado, USA.

Kaback, David B. 2013. "The Modest Beginnings of One Genome Project." *Genetics* 194 (2): 291–99.

Kay, Lily E. 1995. "Who Wrote the Book of Life? Information and the Transformation of Molecular Biology, 1945–55." *Science in Context* 8 (4): 609–34.

Kay, Lily E. 1998. "A Book of Life?: How the Genome Became an Information System and DNA a Language." *Perspectives in Biology and Medicine* 41 (4): 504–28.

Kay, Lily. 2000. *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford: Stanford University Press.

Keller, Evelyn Fox. 2009. “Knowing As Making, Making As Knowing: The Many Lives of Synthetic Biology.” *Biological Theory* 4 (4): 333–39.

Keller, Evelyn Fox. 2011. “Genes, Genomes, and Genomics.” *Biological Theory* 6 (2): 132–40.

Keller, Evelyn Fox. 2012. “Lexicons, Kind-Terms, and World Changes.” *Historical Studies in the Natural Sciences* 42 (5): 527–31.

Keller, Evelyn Fox. 2015. “The Postgenomic Genome.” In *Postgenomics: Perspectives on Biology after the Genome,* edited by Sarah S. Richardson and Hallam Stevens, 9-31. Durham: Duke University Press.

Kohler, Robert E. 1994. *Lords of the fly: Drosophila* genetics and the experimental life. Chicago: University of Chicago Press.

Korhola, Matti. 2018. “Between Science and Industry - Applied Yeast Research.” *FEMS Yeast Research* 18(2): foy008.

Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Langer, Erika Michelle. 2016. “Molecular Ferment: The Rise and Proliferation of Yeast Model Organism Research.” Ph.D., United States -- California: University of California, San Francisco.

Landecker. Hannah. 2007. Culturing life: How cells became technologies. Cambridge: Harvard University Press.

Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Law, John. 1999. “After ANT: Complexity, Naming and Topology.” *The Sociological Review* 47 (S1): 1–14.

Law, John, and Marianne Elisabeth Lien. 2012. “Slippery: Field Notes in Empirical Ontology.” *Social Studies of Science* 43 (3): 363–78.

Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago and London: University of Chicago Press.

Leonelli, Sabina, and Rachel A. Ankeny. 2012. “Re-Thinking Organisms: The Impact of Databases on Model Organism Biology.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 29–36.

Levy, J. 1994. “Sequencing the Yeast Genome: An International Achievement.” *Yeast* 10 (13): 1689–1706.

Lindegren, Carl C. 1949. *The Yeast Cell: Its Genetics and Cytology*. Saint Louis: Educational Publishers, Inc.

Mol, Annemarie. 2002. *The Body Multiple*. Durham: Duke University Press.

Mortimer, Robert K. Lawrence Berkeley Life Sciences Division Technical Records, Undated. Berkeley, California, USA.

Myers, Greg. 1990. *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. Madison: The University of Wisconsin Press.

Myers, Greg 1996. "Out of the Laboratory and Down to the Bay: Writing in Science and Technology Studies." *Written Communication* 13(1): 5-43.

Nelson, Nicole C. 2018. *Model Behavior: Animal Experiments, Complexity, and the Genetics of Psychiatric Disorders*. Chicago: University of Chicago Press.

Nerlich, Brigitte, and Iina Hellsten. 2004. "Genomics: Shifts in Metaphorical Landscape between 2000 and 2003." *New Genetics and Society* 23 (3): 255–68.

Noguera-Solano, Ricardo, Rosaura Ruiz-Gutierrez, and Juan Manuel Rodriguez-Caso. 2013. "Genome: Twisting Stories with DNA." *Endeavour* 37 (4): 213–19.

Parolini, Giuditta. 2018. "Building Human and Industrial Capacity in European Biotechnology: The Yeast Genome Sequencing Project (1989–1996)." 1–43. dx.doi.org/10.14279/depositonce-6693

Parry, Sarah, and John Dupré. 2010. "Introducing Nature after the Genome." *The Sociological Review* 58 (1_suppl): 3–16.

Pollack, Andrew. 2016. "Scientists Announce HGP-Write, Project to Synthesize the Human Genome." *The New York Times*, June 2, 2016, sec. Science.

Richardson, Sarah S., and Hallam Stevens. 2015. "Beyond the Genome" In *Postgenomics: Perspectives on Biology after the Genome,* edited by Sarah S. Richardson and Hallam Stevens, 1-8. Durham: Duke University Press.

Richardson, Sarah M., Leslie A. Mitchell, Giovanni Stracquadanio, Kun Yang, Jessica S. Dymond, James E. DiCarlo, Dongwon Lee, et al. 2017. 'Design of a Synthetic Yeast Genome'. *Science* 355 (6329): 1040–44.

Sapp, Jan. 2003. *Genesis: The Evolution of Biology*. Oxford: Oxford University Press.

Stevens, Hallam. 2013. *Life Out of Sequence: A Data-Driven History of Bioinformatics*. Chicago: University of Chicago Press.

Stevens, Hallam. 2018. "Globalizing Genomics: The Origins of the International Nucleotide Sequence Database Collaboration." *Journal of the History of Biology* 51 (4): 657-691.

Szybalski, Waclaw. 2001. "My Road to Øjvind Winge, the Father of Yeast Genetics." *Genetics* 158 (1): 1–6.

Szymanski, Erika Amethyst. 2018. "Remaking Yeast: Metaphors as Scientific Tools in *Saccharomyces cerevisiae* 2.0." *BioSocieties*, September.

Szymanski, Erika, and Jane Calvert. 2018. “Designing with Living Systems in the Synthetic Yeast Project.” *Nature Communications* 9 (1).

“The Yeast Genetics Community and its Conferences.” 1986. *Yeast*, 2 (S1986A): i-ii.

Thorne, R. S. W. 1947. “Inheritance in Yeast.” *Journal of the Institute of Brewing* 53 (1): 25–36.

Zhu, Yiwen, and Darren Kuang. 1992. “Construction of the Physical Map of Yeast (Saccharomyces Cerevisiae) Chromosome V.” *Cell Research* 2 (1): 25–34.